Jim

Leslie

As a basis for discussion.

Pierre
23 Feb 72

Argument.

1. The amino acid sequence data show that there is no obvious restriction on the sequence of amino acids, except possibly for the rarer ones. It is therefore very likely that the code is non-overlapping.

2. If the code is non-overlapping we have stereochemical difficulties It may be possible to overcome these if we invoke more than one nucleic acid chain, though we then have the problem of bringing together specifically chains with different base sequences.

In this approach we therefore ignore these stereochemical difficulties completely. We can justify this by invoking a mechanism of the type suggested by Sydney Brenner, the essential feature of which is that the completed polypeptide chain is not attached to the template.

3. If the code is non-overlapping it is likely, but not certain, that there are restrictions on the possible base sequences, since we cannot code for just 20 different things without restrictions, except in an arbitary manner. It remains to be seen whether these differences are of an obvious or of a subtle kind.

4. We now consider the experimental data on the turnover of RNA. These show (a) that DNA is not necessary for this

(b) that neither protein synthesis nor amino acid incorporation is necessary either.

I suspect that there is evidence to show that these two points are also true for nett RNA synthesis. (It would be interesting to know if there is a limit to RNA turnover or nett synthesis when protein synthesis is blocked). Thus we must have an RNA - RNA copying mechanism.

5. It is difficult to conceive a simple copying mechanism in which like goes with like (In passing: we should nevertheless spend more time thinking how this might be done) We thus assume an RNA - RNA complementary mechanism. We shall usually assume the DNA type of complementarity.

I think it is true to say that so far we have nothing very new. We now proceed to let these assumptions interact. Thus:

6. Since (from 3) there are restrictions on the base sequence, and since (by 5) we also have the complementary sequence, we must ask : does the complementary sequence make sense ? To discuss this we make the following postulate: a base sequence found in nature either makes complete sense or complete nonsense; an intimate mixture of sense and nonsence is forbidden.

7. We now have two possibilities:

(a) either the complement of a sensible sequence is sense

or (b) it is nonsence.

We choose the latter because

(i) it seems unlikely, because too restrictive, that two different polypeptide chains are made on the same jig, that is from the same stretch of information.

(ii) the insulin sequence data for different species show that it is very unlikely to be the same polypeptide chain turned round i.e. one part on one RNA chain and the other part on its complement.

From all this we therefore deduce:

There are two types of RNA chains, having different restrictios on the sequence of bases. Each type is the complement of the other. One base sequence makes sense everywhere; the other nonsense everywhere.

Protein synthesis (and amino acid incorporation) is linked with the synthesis of only <u>one</u> of the two types of RNA chain.

In order to say anything interesting about the code we now have to introduce another postulate. This could have been introduced as independent of some of the earlier ones, but it is more convenient to do so here.

8. If the code is non-overlapping it is probable (though not certain) that we need some arrangement to enable us to read the correct group of nucleotides. This could be a structural comma (note that a coiled coil structure might provide this) . We arbitarily choose the other alternative: that the restrictions inherent in the code prevent a polynucleotide group from going into the wrong place.

Our work on codes goes in here. We have shown

(a) that, neglecting complementarity, a 3-group code gives just 20 possible 3-groups. (Incidentally we can make a ~~three~~ 3-group code for 16 amino acids the complement of which makes <u>sense</u> everywhere. We can extend this to 20 if we put restrictions on the neighbours of the extra four)

(b) allowing for complementarity, a 4-group code, which makes sense on one chain and nonsense everywhere on the other, certainly gives 21 possible groups: the maximum is so far unknown but 27 is an upper bound.

Notice that we have assumed in all this that we know the <u>direction</u> of reading of a base sequence. This is stereochemically reasonable but will eventually need stereochemical justification. Another postulate for possible codes is that they make nonsense everywhere <u>backwards</u>, either with or without complementarity.

That completes the formal framework of our thinking and we can now proceed to develope particular schemes within it.

Synthesis of DNA

The usual scheme, but varients are possible, in that the building units may be polynucleotides rather than mononucleotides. The data* hints that the latter is unlikely (*the incorporation of unusual bases). Perhaps its desirable to make at least four hydrogen bonds at an attachment. Alternatively, if the language is redundent, it may make mistakes less frequent if polynucleotides are used.

The synthesis of RNA on DNA

The previous scheme had two formal difficulties (apart from the stereochemical one)

(a) only one hydrogen bond was formed per base

(b) there was nothing to prevent the complementary reading of the code.

We now easily overcome these by using our complementary 4-group code. This sets us a stereocheical problem: can it be built ? can a plausible reason be given to show how the direction of reading the code is decided? It is worth remembering that protamine stabilises the structure in form B. Perhaps form A is the active form, and form B the resting form .

We have a choice as to whether this RNA synthesis is coupled with protein synthesis or not.

RNA and protein synthesis

Several schemes appear possible. Let the two types of RNA chain be called P and Q (P linked with protein synthesis)

Scheme I

(a) starting with a P chain of RNA. By DNA-type pairing we produce a Q chain. This could be from mononucleotides (plus Ochoa's enzyme) or could be from special polynucleotides, in neither case linked with amino acids.

(b) starting with a Q chain of RNA. By DNA-type pairing we produce a new P chain plus a polypeptide chain, using 4-groups plus amino acids as precursors. It is more reasonable to assume that this is coupled obligitorily with protein synthesis, but this may not be essential.

Scheme II

As scheme I, but we add another possible synthesis, namely one like the RNA-on-DNA one. That is

(c) we synthesise a P chain, plus a polypeptide, on a PQ pair, using 4-groups plus amino acids.

Scheme III

We can also make a scheme with processes (a) and (c) only. This is rather elegant. We have a choice as to whether we unwind the paired chains or not.

Some general remarks

1. One of the best authenticated experimental results is that if one amino acid is cut off or blocked we stop the use of all other amino acids, apart from Gale's incorporation. It seems to me that we should try to think what this implies.

2. If all these ideas are sound, what is the RNA structure we see with X-rays ? Is it just an artefact, or can it be fitted into the scheme ?

3. Our present schemes seem to produce too much nett RNA synthesis. Can we produce any good ideas about the breakdown of RNA ?